Colaço *et al.*
Large-scale Assessments in Education (2025) 13:1
https://doi.org/10.1186/s40536-024-00234-7

Large-scale Assessments in Education

RESEARCH **Open Access**

Understanding the private-public school performance gap in PISA: evidence from Portugal

Ricardo Colaço[1], Pedro Freitas[2,3], Luis Catela Nunes[1*] and Ana Balcão Reis[1]

*Correspondence: lcnunes@novasbe.pt

[1] Nova School of Business and Economics, Universidade Nova de Lisboa, R. da Holanda 1, 2775–405 Carcavelos, Portugal
[2] Blavatnik School of Government, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG Oxford, United Kingdom
[3] Nova SBE Economics of Education Knowledge Center, Universidade Nova de Lisboa, Carcavelos, Portugal

Abstract

We analyse the PISA-reported convergence in the performance of private and public schools in Portugal. When PISA sampling weights are used, the number of students enrolled in those types of schools and specific grades/tracks of study differs significantly from official population figures. To account for those differences, we apply a post-stratification adjustment; however, sample sizes are small, resulting in estimates with low precision for several subgroups. We propose recommendations for improving the handling of these issues in future PISA cycles. In an additional analysis, we also account for changes in the composition of the student population. When all factors are considered, the convergence in scores is far less impressive than reported. For instance, in Science, after adjusting the sampling weights and removing population composition effects, the reported convergence of 46 points between private and public schools from 2015 to 2018 amounts to only 9 points. The decomposition and sample adjustment methods used in this paper can be easily adapted to other contexts.

Introduction

International large-scale assessments have played a major role in the education policy debate in recent years. Among these is OECD's triennial Programme for International Student Assessment (PISA) assessing student's literacy in Reading, Mathematics, and Science. The PISA survey is constructed in such a way as to be representative of the 15-year-old student population at the year the tests are run, comprising students from different regions and types of schools. To ensure that all parts of a population are included in the sample, it follows a stratified sampling methodology based on explicit and implicit stratification variables specifically chosen for each country. The results from PISA have been used to study many different questions such as the sources of inequality in student performance (Woessmann 2016), the effect of the variation in instruction time on students' results (Lavy 2010), how different evaluation methods across countries explain results differentials among participating countries (Bergbauer et al. 2018), or the magnitude of peer effects (Vandenberghe and Robin 2004).

In this work, we focus on the comparison between public and private schools. According to PISA 2018, on average, 18% of the students were enrolled in private schools, a percentage that varies substantially across different education systems. Several studies have used data from PISA to compare the performance of different types of schools, such as Mancebón et al. (2012), Delprato and Chudgar (2018) or OECD (2020).

The pertinence of this topic is also highlighted by the fact that in most of the PISA participating countries, the type of school was one of the chosen dimensions for stratification: of the 82 countries participating in PISA 2018, the public or private nature of the schools was considered as an explicit stratification variable in 33 countries and as an implicit stratification variable in 16 other countries.

The composition of the PISA samples has been a topic of discussion in a number of studies. O'Leary (2001) compares the results of an age based assessment, like PISA, with other assessments that target specific grades. Other studies report several issues regarding PISA samples, such as in the cases of Austria, where in the year 2000 the students in vocational programmes were incorrectly weighted (Neuwirth 2006), in England and the UK (Jerrim 2013, 2021; Durrant and Schnepf 2018), in Sweden where the criteria for student exclusion from the sample were not followed (Andersson and Massih 2023) and Canada where the missing observations and the non-response rates of schools and students jeopardized the representativeness of the PISA results (Anders et al. 2020). For Portugal, Freitas et al. (2016) compared the distribution of the PISA data between 2006 and 2012 in terms of school grade and track of studies finding substantial differences between the sample and the population.

We also consider the case of Portugal and analyse the recently reported convergence in the performance of private and public schools. When PISA sampling weights are used, the number of students enrolled in those types of schools and specific grades/tracks of study differs significantly from official population figures. To account for those differences, we apply a post-stratification adjustment that makes use of official population data to adjust the students sampling weights employed in the calculation of the point estimates. Although sample sizes are small and estimates have low precision for several subgroups, we proceed with this analysis and suggest how to deal with these issues in future PISA cycles.

In an additional analysis, we also account for changes in the composition of the student population. The evolution of average scores is decomposed into changes in student performance and changes in the composition of the student population enrolled in each type of school. As we will discuss, we find marked changes in the composition of the student population in private schools, where the percentage of students in vocational tracks has been increasing throughout the years.

When all factors are considered, the convergence in scores is far less impressive than reported. For instance, in Science, after adjusting the sampling weights and removing population composition effects, the reported convergence of 46 points between private and public schools from 2015 to 2018 amounts to only 9 points.

Although the small sample size of private schools in PISA leads to poor precision estimates of the private-public gap and large confidence intervals, which our post-stratification exercise cannot address, the approach followed in this work can be easily applied to any other participating PISA countries to calculate adjusted sampling weights, generate adjusted scores, and derive the decomposition of the variation in scores over time.

The rest of this paper is organized as follows. "PISA and Population Data" section analyses the PISA and population data, "Methods and results" section develops on the methods and results that account for differences between the PISA sample and the student population distribution, and "Conclusion " section concludes.

PISA and population data

Between 2000 and 2015, Portugal was one of the countries with the largest improvements in PISA scores, having the third largest increase in Reading and the highest increase in Mathematics and Science. In the 2018 PISA wave, the results plateaued compared with previous waves. And when the data is decomposed between public and private[1] schools, we observe large falls in the results of the students enrolled in private schools, with a negative variation of 43 points in Reading, 47 in Mathematics and 56 in Science. This variation comes after a sustained improvement in the results of private schools students in the previous PISA waves, driving the 2018 results back to the 2006 level. Such a sharp decline raised concerns in the media regarding the sub-sample of private schools in the 2015 and 2018 waves making it an interesting case for further study.[2]

The PISA survey aims to be a representative sample of the 15-year-old student population, the target population of the PISA test. In the vast majority of the participating countries, the sampling follows a two-stage design. The first-stage consists of sampling units of individual schools with 15-year-old students. The second-stage sampling units are students within the sampled schools.

The schools participating in PISA are chosen according to a school stratification process, based on explicit and implicit stratification variables (OECD 2017). The explicit stratification divides the country's schools into different strata within which the participating schools are selected. In Portugal, the geographic region was used in both 2015 and 2018. The modal grade was also used as an explicit stratification variable in 2015. The implicit stratification consists of sorting schools within each explicit stratum according to the implicit stratification variables, ensuring a proportional representation of schools across the chosen dimensions. For Portugal, school funding (whether funded by the government or not) or the type of curriculum offered (academic or vocational oriented schools) were used as implicit stratification variables in both 2015 and 2018.[3]

For sampled schools a response rate of 85% was required. In case the response rate is between 65% and 85% replacement schools previously listed could be used. However, for countries where the participation rate has not increased after using replacement schools, a further non-response bias analysis is conducted. Regarding student participation rate, an overall level of 80% is required, and for each school a participation rate of 50% is necessary. In Table 7 in Appendix 1, we detail the response rates at the school and student level in 2015 and 2018. In 2018 the student response rate is below the 80% threshold, triggering a further analysis by the OECD on the validity of the sample. Data on response rates by type of schools, public or private, and the original sampling used

[1] In the definition of private schools, we include privately owned schools that receive public funding.

[2] https://www.publico.pt/2020/01/20/sociedade/opiniao/amostra-pisa-errada-realidade-alterou-1,899,887 or https://observador.pt/opiniao/pisa-2018-e-o-erro-na-amostra/

[3] A summary of the explicit and implicit stratification variables is presented in Table 8 in Appendix 2.

by OECD are not publicly available, which does not enable us to observe which type of schools depicted a higher prevalence of non-response.

Additionally, PISA includes survey weights to guarantee that each sampled student represents the adequate number of students in the target population. PISA weights are constructed using two base weights, the school base weight and the within school base weight which are adjusted considering: 1. non-participation by other similar schools, incorporating the explicit and implicit dimensions considered in the stratification; 2. cases when observed 15-year-old students were enrolled in the PISA modal grade (not used in 2018); 3. non-participation by students within the school; 4. trimming of large school weights; 5. trimming of large student weights. As pointed by Anders et al. (2020), the effectiveness of this weighting correction method for non-response is jeopardized if there are some other dimensions which are both predictive of non-participation and correlated with PISA scores. If that is the case, the method would assume that the non-response by schools and students within PISA occurs at random, which may not be the case as previously shown by Micklewright et al. (2012) or Heine et al. (2017).

We make use of two different data sources: PISA data from 2015 and 2018 provided by OECD and Portuguese population data registered in the administrative microdata on the Portuguese educational system provided by *Direção Geral de Estatisticas da Educação e Ciência (DGEEC)*, which follows all the students enrolled in schools in mainland Portugal between 2007 and 2018.[4] We start by examining the PISA sample according to the distribution of students enrolled in private and public schools.[5] As shown in Table 1, in 2015, the share of 15-year-old students enrolled in private schools according to the PISA sampling weights, 6.6%, is meaningfully lower than the one in the population data, 14.1%. This substantial imbalance is not present in the following wave in 2018.

In Portugal, the PISA samples include students in the lower secondary education (Lower SE), which comprises the 7th, 8th, and 9th grades, and in the upper secondary education (Upper SE), which includes the 10th and 11th grades[6] and in academic and vocational tracks.[7] Based on this structure we detail in Table 2 the corresponding shares of students according to PISA and the population datasets in 2015 and 2018. In these subgroups, we focus on dimensions assumed in the implicit stratification of the sample, whose goal is to ensure the strictly proportional sample allocation across the groups considered. In the Portuguese case, grade and curriculum or programme orientation and school funding were considered as implicit stratification variables both in 2015 and 2018 (Table 8 in Appendix 2).

In 2015, we find large differences between the PISA sample and the population. The sample included a smaller proportion of students in the 7th and 8th grades and in vocational tracks, particularly at the lower secondary level. This was compensated by a higher

[4] For proper comparability, the data regarding the autonomous regions of Azores and Madeira in PISA were excluded.

[5] We label the students as enrolled in public or private schools according to the question in the school questionnaire "Is your school a public or a private school?" (SC013). In 2015, four schools did not answer this question. In 2018, for two schools the answer to this question was not consistent with the variable PRIVATESCH, an additional PISA variable that labels the schools as private or public. To fill these information gaps in PISA data, we cross it with the final reports on the PISA tests published by the Portuguese Ministry of Education.

[6] We merged the students in the 10th and 11th grade given the low weight that the 11th grade students have in the total population. The 10th grade corresponds to the modal grade of the PISA data in Portugal.

[7] The artistic track was excluded from the analysis, since its curriculum differs from the traditional academic and vocational tracks and it amounts to only 0.6% of the total student population, in both 2015 and 2018.

Table 1 Share of Students in Private Schools

	2015	2018
Population	14.1%	13.1%
PISA (final student weights)	6.6%	14.9%
	(1.09)	(1.76)

Standard errors based on replicate weights in parentheses

proportion of students in the academic track at the upper secondary level. In 2018, the differences were much smaller, particularly in the vocational tracks; however, in the academic track, the PISA sample showed a smaller share of lower secondary students and a higher share of upper secondary students.

The split of the data between the students enrolled in public and private schools shows significant disparities in both waves. For public schools, the differences between the PISA sample and the official population are mostly in line with the differences observed above for the overall student population. For private schools the picture is substantially different. In 2015 we notice a larger share of students enrolled in the academic track of the upper secondary in the PISA sample – 13 percentage points more students in these grades and tracks when compared to the population. This is made up for by lower shares of students in the lower secondary, particularly of those in the vocational track. In 2018 these differences reverse, since the students in the vocational tracks present a higher share in the PISA sample, which is offset by a lower share in all the other grades/tracks, especially of those in the 9th grade. Several differences are large in magnitude, particularly for private schools in 2015. However, some of them are not statistically significant, which is explained by the large standard errors for some groups like the academic upper-secondary. These large standard errors in 2015 are also explained by the limited sample sizes for private schools in terms of number of students and schools in that year, as shown in Table 3.

The observed differences between the sample and population distributions across grade, track of studies and type of school can emerge due to several reasons, from a defective initial sampling frame to the potential inability of the weighting adjustments to capture non-response patterns across schools and students within schools. Since schools are sampled one year before the tests take place, and since this sampling is based on data from the year before that, it could also be that these differences are explained by a change in the population structure from 2 years earlier. However, as can be seen in Appendix 3 (Table 9), this is not the case for Portugal. As the OECD does not disclose data on the initial sampling frame or the non-response patterns, we can not fully identify the source behind such sizeable gaps.

To understand the relevance of these differences for the overall results, it is important to analyse the PISA scores, measured as the mean of the estimated score distribution, for each subgroup of students.[8] In Table 4 we can see that the mean score for the modal subgroup (Upper SE) is significantly higher than the mean score for any other subgroup.

[8] A single point estimate of the student's latent ability may be a noisy measure, and that is why OECD reports per student a range of proficiency measures–the plausible values. For purposes of measuring the average PISA result, for both the whole sample and the different subgroups, the 10 plausible values are averaged out, OECD (2009).

Table 2 Distribution PISA/Population students

Grade/Track of studies	2015			2018		
	Population	PISA	Difference	Population	PISA	Difference
All schools						
7th grade (Academic)	3.80%	2.65% (0.31)	− 1.15[c]	3.41%	2.24% (0.26)	− 1.17[c]
8th grade (Academic)	8.12%	7.04% (0.40)	− 1.08[c]	8.17%	6.98% (0.46)	− 1.19[c]
9th grade (Academic)	18.90%	19.88% (0.88)	0.98	17.74%	17.30% (0.96)	− 0.43
Upper SE (Academic)	52.49%	57.18% (1.38)	4.69[c]	55.15%	57.50% (1.42)	2.35a
Lower SE (Vocational)	6.88%	4.60% (0.41)	− 2.28[c]	3.53%	3.86% (0.87)	0.33
Upper SE (Vocational)	9.81%	8.65% (1.17)	− 1.16	12.00%	12.11% (1.22)	0.12
Public schools						
7th grade (Academic)	4.19%	2.84% (0.33)	− 1.35[c]	3.82%	2.56% (0.30)	− 1.26[c]
8th grade (Academic)	8.76%	7.53% (0.40)	− 1.23[c]	9.04%	7.97% (0.52)	− 1.08[b]
9th grade (Academic)	20.15%	20.81% (0.96)	0.66	19.02%	19.46% (1.00)	0.44
Upper SE (Academic)	52.56%	56.61% (1.43)	4.06[c]	56.36%	59.83% (1.31)	3.47[c]
Lower SE (Vocational)	7.03%	4.92% (0.43)	− 2.11[c]	2.88%	2.36% (0.50)	− 0.52
Upper SE (Vocational)	7.30%	7.28% (0.76)	−0.02	8.89%	7.83% (0.76)	−1.06
Private schools						
7th grade (Academic)	1.44%	0.00%	−1.44	0.68%	0.39% (0.19)	−0.3
8th grade (Academic)	4.20%	0.00%	−4.2	2.39%	1.36% (0.49)	−1.03[b]
9th grade (Academic)	11.30%	6.69% (4.71)	−4.61	9.20%	5.00% (1.19)	−4.2[c]
Upper SE (Academic)	52.07%	65.25% (10.89)	13.18	47.16%	44.23% (7.07)	−2.93
Lower SE (Vocational)	5.99%	0.00%	−5.99	7.88%	12.43% (4.65)	4.55
Upper SE (Vocational)	24.99%	28.06%	3.07	32.69%	36.59% (6.22)	3.9

Statistically significant at [a]10%, [b]5% and [c]1% for one sample z-test of proportions. Standard errors for the PISA sample proportions are shown in parentheses next to the proportions (for private schools in 2015, three subgroups were not sampled). For each combination of year (2015, 2018) and type of school (All, Public, Private) we additionally performed chi-squared goodness-of-fit tests adjusted for survey design. Chi-squared statistics (p-values in parentheses): 2015 All schools: 104.06 (0.00); 2015 Private schools: 52.02 (0.18); 2015 Public schools: 83.53 (0.00); 2018 All schools: 39.86 (0.00); 2018 Private schools: 43.55 (0.24); 2018 Public schools: 47.15 (0.06). Chi-square statistics calculated using the mgof STATA command, Jann (2008).

One can also see that the overall mean score of private schools fell significantly in 2018 for the three domains. The mean scores for public schools remained relatively stable. However, when analysing by subgroups[9] of grades and track of studies, one can see that while the scores for the vocational track of the upper secondary education in private schools indeed fell, that is less prevalent in the case of academic track of the upper secondary education. This illustrates how the comparison between private and public schools as a whole can be misleading, as well as the importance of the distribution of the sample across grades and tracks of study.

Framing this data in the OECD context, in Appendix 5 and 7 we show the percentage of students attending a private school[10] and the percentage of students who are on a vocational track,[11] according to PISA data. It is apparent that there is considerable cross-country variation in both of these dimensions, as well as some within-country variation although to a lesser extent. If we consider the median percentage of students in private

[9] In Appendix 4 we detail how scores were computed for the subgroups that do not have students sampled in PISA 2015.

[10] In Table 11 in Appendix 6, we also show the percentage for which this information is missing.

[11] This classification was made by the authors, using the variable PROGN, by searching for strings that included "vocational" and "technical" as well as manual verification for possible exceptions. It is thus suggestive of the percentage of students in vocational tracks in PISA, but for some country-years there is no such distinction so these statistics must be looked at with caution.

Table 3 Number of students and schools in the PISA samples

Grade/Track of studies	2015		2018	
	Public	Private	Public	Private
	# Students/Schools	# Students/Schools	# Students/Schools	# Students/Schools
7th Grade (Academic)	149/74	0/0	98/66	3/2
8th Grade (Academic)	422/134	0/0	342/138	11/7
9th Grade (Academic)	1181/166	14/4	814/191	50/13
Upper SE (Academic)	2886/131	218/6	3023/178	361/18
Lower SE (Vocational)	284/91	0/0	119/45	66/12
Upper SE (Vocational)	384/94	67/3	397/123	265/26
Total	5306/690	299/13	4793/741	756/78

schools by country, we can see that 16 countries have more than 10% of students in private schools.

Since attending a private school is generally associated with higher socio-economic status, this is an important dimension to consider. If we consider the median percentage of students in a vocational track by country, there are 17 countries which have more than 10% of students in vocational tracks. Since students in vocational tracks tend to come from a lower socio-economic status and to have a lower achievement on average, this is another relevant dimension.

Methods and results

To account for the differences found between the population and the PISA sample weights, we use adjusted sampling weights, as in Freitas et al. (2016).

PISA scores aim to measure the latent ability of a student in the Reading, Maths and Science dimensions.[12] We take PISA scores as valid for each group of students and adjust the weight of each group according to population data.

Considering the division according to subgroups of grade and track of studies, the adjusted PISA score is computed as:

$$PISA_{adjustedscore} = \sum_{j=1}^{J} w^j \overline{PISA}^j_{score} \tag{1}$$

where w^j is the share of students of subgroup j in the population according to the population data and $\overline{PISA}^j_{score}$ is the average PISA score for the same subgroup j.[13] Each subgroup j is taken as the grade and track of studies in private or public schools considered in Table 2. For the adjusted scores we compute the respective standard errors multiplying each individual weight reported in the PISA database by the adjustment factor that

[12] PISA scores calculated using the command repest in STATA, Avvisati & Keslair (2016).

[13] Population can stand for the overall population or for the population of public or private schools. For the overall population each subgroup j corresponds to grade, track of studies, and school type.

Table 4 PISA Scores - Public and Private Schools

Grade/Track of studies	2015			2018		
	Reading	**Maths**	**Science**	**Reading**	**Maths**	**Science**
Public schools						
7th grade (Academic)	372	367	377	358	358	362
	(7.1)	(7.0)	(6.0)	(9.0)	(9.2)	(7.6)
8th grade (Academic)	408	396	414	383	382	396
	(5.0)	(5.5)	(4.1)	(4.3)	(4.9)	(4.3)
9th grade (Academic)	450	435	451	441	432	441
	(3.9)	(3.3)	(3.0)	(4.1)	(4.8)	(4.2)
Upper SE (Academic)	546	541	548	540	543	537
	(2.8)	(2.5)	(2.4)	(1.8)	(2.1)	(2.1)
Lower SE (Vocational)	377	365	383	346	354	357
	(6.4)	(7.0)	(4.6)	(6.7)	(8.0)	(7.1)
Upper SE (Vocational)	472	467	475	463	467	459
	(6.8)	(7.0)	(6.2)	(4.5)	(4.8)	(4.7)
Overall	497	489	499	493	493	492
	(2.9)	(2.8)	(2.6)	(2.5)	(2.5)	(2.5)
Private schools						
7th grade (Academic)	377[a]	387[a]	394[a]	383	393	421
	–	–	–	(19.5)	(31.4)	(20.9)
8th grade (Academic)	413[a]	418[a]	433[a]	374	391	403
	–	–	–	(14.4)	(23.5)	(19.3)
9th grade (Academic)	459	438	474	440	452	450
	(16.3)	(13.0)	(9.2)	(21.8)	(18.9)	(15.4)
Upper SE (Academic)	553	570	573	554	566	560
	(12.5)	(7.6)	(11.7)	(7.5)	(8.7)	(8.4)
Lower SE (Vocational)	401[a]	397[a]	406[a]	364	364	361
	–	–	–	(5.9)	(10.1)	(13.7)
Upper SE (Vocational)	503	508	503	470	469	468
	(21.4)	(23.4)	(18.8)	(7.5)	(7.7)	(6.2)
Overall	533	544	547	491	497	494
	(11.0)	(9.1)	(10.5)	(11.0)	(12.5)	(13.3)

Standard errors in parentheses, following the methodology proposed by PISA. [a]Imputed values (see details in Appendix 4). Standard errors not reported for imputed values. All (survey design-adjusted) z-tests of differences between average score for the modal subgroup (Upper SE) and each of the other subgroups are significant at the 5% level (highest p-value is 0.0187, all others are below 0.01)

makes the share of students represented in each subgroup match the share reported in the population data. The standard errors are computed according to the methodology proposed by PISA using these adjusted weights.[14]

The post-stratification exercise results in low precision estimates due to the large standard errors of the PISA scores for subgroups with smaller sample sizes, such as the students in private schools, especially in 2015. As a result, there may be an overlap between the confidence intervals for the updated point estimates and the original confidence intervals as reported by the OECD.

[14] Under the approach proposed by the OECD, the final variance is the combination of the sampling and imputation variances which take into account the different replicate weights and plausible values, OECD (2009).

Table 5 shows the official PISA scores and the recalculated PISA scores according to (1). In 2015 in all the three fields, Reading, Mathematics, and Science, the adjusted PISA scores are 5 to 7 points lower than the official ones, both for the overall data and for the students enrolled in the public schools. However, the difference between the official and adjusted scores for the private schools data is significantly wider, with a difference between 20 and 24 points in Reading, Maths, and Science. This result is expected given that for 2015 we observe in private schools an overrepresentation in PISA of the students in the academic secondary education whose performance is, on average, higher than of those who are underrepresented, namely the ones enrolled in the vocational tracks.

In 2018 the differences between adjusted and official scores are lower, between 4 and 5 points for the set of all students and for the ones in public schools. For the case of the students in private schools, the adjusted scores are higher than the official PISA scores by 5 to 6 points. This goes in line with the greater alignment between the PISA sample distribution and the population data for the private schools.

Merging the results obtained in Freitas et al. (2016) for 2009 and 2012 with the ones derived for the years 2015 and 2018 we plot in Fig. 1 the evolution of the PISA scores between 2009 and 2018 for all the students and separately for public and private schools.[15] We show the adjusted scores and the 95% PISA confidence intervals for the official scores based on the PISA standard errors in order to show how large are the differences between adjusted scores and the official ones published by OECD.

Looking at the evolution of the PISA scores from 2009 and for the students enrolled in public schools, the adjusted scores for the three domains are usually lower than the official ones throughout the years, with the exception of 2012. Generally the trend of the adjusted scores follows that of the official ones, but depicts a smoother evolution across time, and in 2009, 2015 and 2018 the adjusted scores for public schools are closer to the lower bound of the 95% confidence interval reported by PISA. For private schools, across all the three domains, in 2009 the adjusted scores are close to the lower bound of the 95% confidence interval. In 2012 all the adjusted scores are below the 95% confidence interval, while in 2015 the adjusted scores are below this same interval in Maths and close to the lower bound in Reading and Science. In the last observed year, 2018, the adjusted score is much closer to the official one published by PISA and within the confidence interval. Thus, the sizeable fall in the 2018 PISA scores in private schools is not due to the difference between the sample and population distributions in this year, but results from the large positive difference between the official and adjusted scores for this same subgroup in the two previous PISA waves, especially in 2015. However, a fall in the private schools results is still observed, though it is 50% to 60% lower than the one reported in the official PISA data. In 2015, the average difference between public and private schools across the three tests in the official PISA scores is around −46 points, which is reduced by one third to −31 points in the adjusted scores.

[15] We have not considered the year of 2006, since the official data used to compare with the PISA data is different from the population data used in the subsequent years. Additionally the structure of the upper secondary system in the vocational track changed substantially between 2006 and 2009.

Table 5 Official and Adjusted PISA Scores

	All Schools		Public Schools		Private Schools	
	2015	**2018**	**2015**	**2018**	**2015**	**2018**
Reading						
Official	499	493	497	493	533	491
	(2.8)	(2.6)	(2.9)	(2.5)	(11.0)	(11.0)
Adjusted	493	489	490	488	512	496
	–	(2.6)	(3.0)	(2.6)	–	(9.7)
Maths						
Official	493	494	489	493	544	497
	(2.6)	(2.8)	(2.8)	(2.5)	(9.1)	(12.5)
Adjusted	487	490	482	488	520	503
	–	(2.8)	(2.9)	(2.6)	–	(11.2)
Science						
Official	502	493	499	492	547	494
	(2.5)	(2.9)	(2.6)	(2.5)	(10.5)	(13.3)
Adjusted	497	489	492	487	526	500
	–	(2.8)	(2.7)	(2.6)	–	(11.5)

Standard errors in parentheses, following the methodology proposed by PISA. As shown in Table 4, for the private schools in 2015 the results were imputed for some subgroups. Thus, standard errors are not reported for private schools and all schools in 2015

Private-public school gaps

The change in the scores of both public and private schools after applying the adjusted weights raises the question of how the gap between the students in the two types of schools has evolved over time. In Fig. 2 we note that, when considering the adjusted PISA scores, the gap between the public and private schools closed substantially between 2009 and 2018. It fell from 29 to 8 points in Reading, from 31 to 15 points in Mathematics and from 25 to 12 points in Science. The evolution of the gap in the adjusted scores is smoother than the one in the official scores reported by OECD, as in the adjusted scores we do not observe the large increase between 2009 and 2012 followed by a large fall between 2015 and 2018.

As seen before, there are differences in the evolution of scores of the academic and vocational tracks. In Fig. 3 we compare the evolution of the private-public gap in the vocational and academic track. The gap in the academic track is considerably more stable than in the vocational one, namely in Mathematics and Science, where the gap remains steady at around 40 points between 2009 and 2018. The private-public school gap on the vocational track peaks in 2015 and then sharply decrease in 2018. For all the three dimensions the private-public school gaps for the vocational track students is between 10 and 15 points in 2018.[16]

Decomposition of the adjusted PISA scores evolution

The evolution observed for students of both private and public schools raises the question of what are the drivers behind the evolution in PISA scores. Namely,

[16] This analysis of the gaps between students in public and private schools by track of studies is based on point estimates whose precision may be low given the smaller sample sizes.

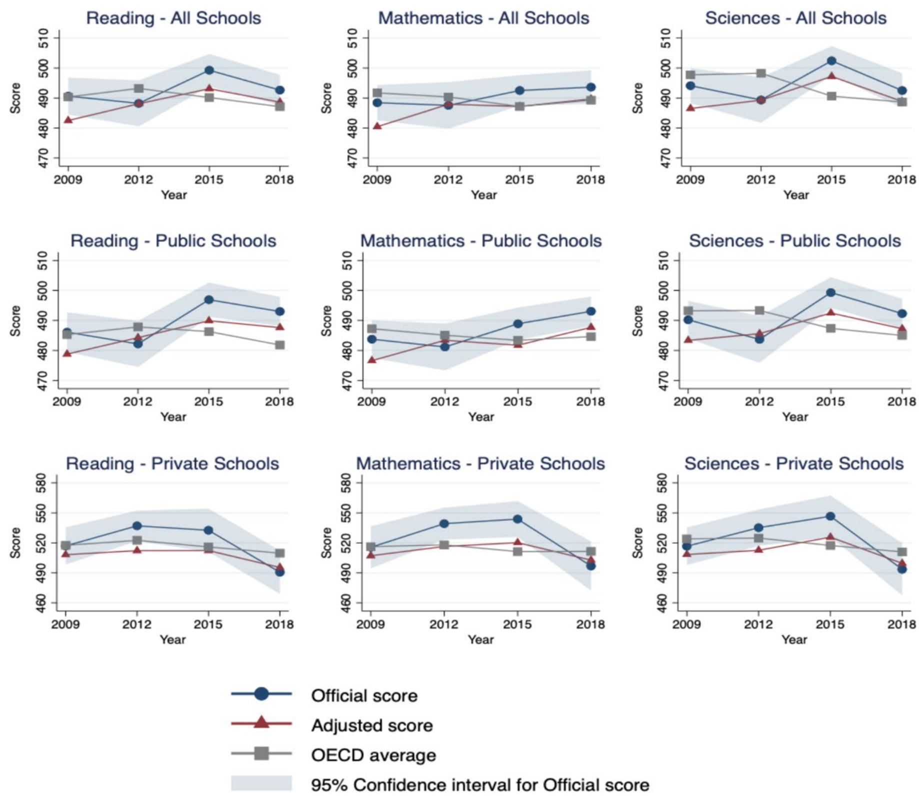

Fig. 1 Adjusted and Official PISA Scores

whether this evolution stems from the change in the structure of the student population or due to the change in the student proficiency in PISA.

Given the structure of the student population across the different PISA waves, we decompose the evolution of the adjusted PISA scores from 2015 to 2018 into two different components: *(i)* the part of the evolution that is due to changes in the population structure, namely in the distribution of students across different grades, tracks of studies and school types; *(ii)* changes due to the evolution in student performance in each of those subgroups. The proposed decomposition follows the intuition behind the Oaxaca-Blinder decomposition ((Oaxaca 1973) and (Blinder 1973)), which in a linear regression model decomposes the differences in the mean of the dependent variable between differences in characteristics and in parameters. This decomposition approach was used before in the context of PISA data to measure the drivers of gender gap in students' achievement (Gevrek et al. 2020) and the reasons for the difference between two high achieving Asian countries, South Korea and Singapore, and a lower achiever in PISA, Malaysia (Perera and Asadullah 2019). In our case, contrary to the traditional decomposition, we do not consider a set of explanatory factors for the PISA scores, but assume only three dimensions—grade, track of studies, and whether the students are enrolled in public or private schools.

Thus, the change in PISA scores between time $t-1$ and t considering the subgroups of students can be decomposed in score and population effects. The score effect assesses

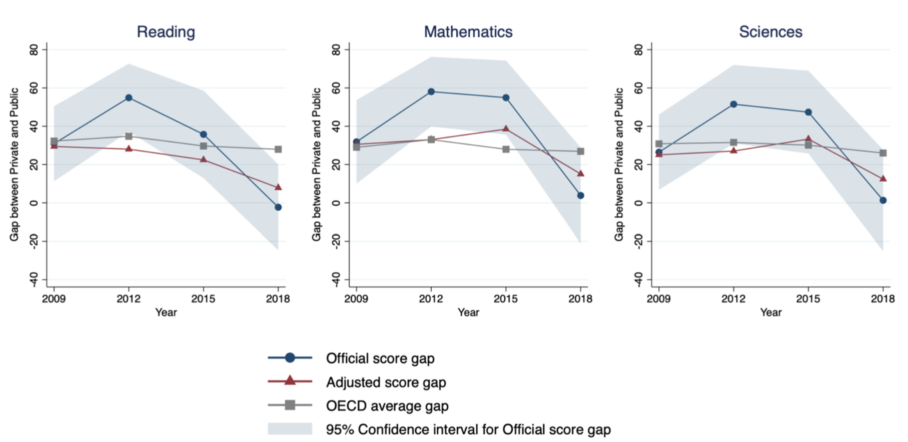

Fig. 2 Private-Public schools gap (2009–2018)

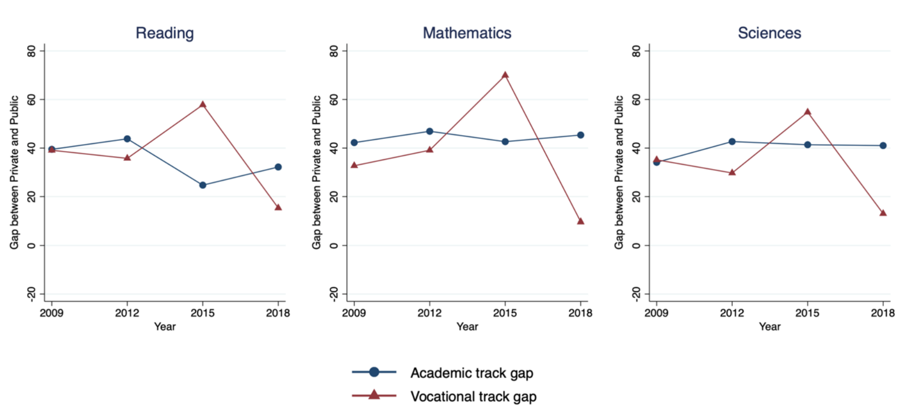

Fig. 3 Private-Public schools gap by track of studies (2009–2018)

the evolution between $t-1$ and t keeping the population composition constant at time $t-1$. It is calculated as the difference between the adjusted PISA scores considering the average performance of each subgroup j at t but the population distribution at $t-1$, and the PISA score at $t-1$. The population effect aims to isolate the contribution of the change in the population composition to the evolution in PISA scores. It is computed as the difference between the adjusted PISA score at time t and the adjusted PISA score at time t but considering the population distribution at time $t-1$. We then consider that the evolution between two time periods, $t-1$ and t is given as:

$$\begin{aligned} PISA_{score}^{t} &= PISA_{score}^{t-1} + \text{Score Effect} + \text{Population Effect} \\ \text{Score Effect} &= \sum_{j=1}^{J} w^{j,t-1}\left(\overline{PISA}_{score}^{j,t} - \overline{PISA}_{score}^{j,t-1}\right) \\ \text{Population Effect} &= \sum_{j=1}^{J} \left(w^{j,t} - w^{j,t-1}\right)\overline{PISA}_{score}^{j,t} \end{aligned} \tag{2}$$

Table 6 reports the results of this decomposition for the overall set of students and for those enrolled in public and private schools. To read the long-run trend of the data we add the results of the same decomposition that are reported for the period between 2009 and 2012 in Freitas et al. (2016).

For the overall sample the score effect dominates as the driver in PISA results between 2009 and 2012, explaining 4 of the 6 point evolution observed in Reading and 3 of the 5 points in Maths. Between 2012 and 2015 the importance of the score effects shrinks substantially, becoming negative for Maths. The contribution of the change in population is positive, but below 5 points. Between 2015 and 2018, the contribution of the population effect increases to around 5–7 points. Also between the last two PISA waves the score effects are negative across the three PISA domains, namely − 11 points in Reading and − 14 in Science. Summing up, for the total sample, the population effect is always positive and is increasing throughout the years. As for the score effect the evolution is not as straightforward, starting with positive to near zero values, and reaching negative values across all domains between 2015 and 2018.

The evolution between score and population effects is different when we consider the public or the private schools. Given the weight of the students enrolled in public schools in the total student population, the results of the decomposition for these students are similar to those reported for the overall sample.

Regarding the decomposition results for the students in private schools between 2009 and 2012 the score effect is the main determinant of the evolution in this period. The score effect becomes negative for Reading and Mathematics between 2012 and 2015 and from 2015 to 2018 we estimate a large negative effect on the score side, between − 22 points in Science and − 13 points in Reading and Mathematics. This patterns follows the one observed in the overall and public school samples. The population effect contributes negatively for the score evolution between 2009 and 2012 and between 2015 and 2018. This differs from the population effect observed in the public and overall samples, which remains positive between 2015 and 2018. This difference is justified by the significant rise for the students in the vocational track (whose performance is, on average, lower) in private schools.

All the results show a common trend of decreasing contributions from the score effect, which has been partially offset by positive population effects, notably in public schools. We also note that the evolution in the Science domain is not always aligned with the change in the other two domains, Reading and Mathematics.

Combining the population effect with the adjustment of sampling weights, we note that the performance gap between private and public schools between 2015 and 2018 still grows smaller, but by much less than what the official results suggest. Taking Science as an example: for private schools, the fall in official scores was of 53 points; but when using adjusted sampling weights the fall shrinks to 26 points; and upon removing the population effect we are left with a reduction of 22 points. For public schools, one goes from a fall of 7, to 5 in the adjusted scores, and to 13 upon removing the population effect. This means that while originally the official scores suggested a convergence of 46 points between private and public schools, when adjusting the sampling weights and removing the population composition effect, they converge by only 9 points.

Table 6 Decomposition of the Variation in Adjusted Scores

	Score effect			Population effect			Total variation		
	Reading	Maths	Science	Reading	Maths	Science	Reading	Maths	Science
All schools									
2009–2012	4.0	3.3	−0.9	1.6	1.9	1.4	5.6	5.2	0.5
2012–2015	0.3	−5.7	3.3	4.6	5.0	4.6	5.0	−0.7	7.9
2015–2018	−11.0	−3.7	−14.1	6.5	6.2	5.7	−4.4	2.5	−8.3
Public schools									
2009–2012	1.1	2.2	−1.9	4.4	4.6	4.1	5.5	6.8	2.2
2012–2015	1.4	−6.0	2.8	4.2	4.4	4.1	5.6	−1.7	6.9
2015–2018	−10.5	−2.2	−12.8	8.2	8.1	7.6	−2.3	6.0	−5.2
Private schools									
2009–2012	6.2	10.8	6.7	−2.1	−1.4	−2.6	4.1	9.4	4.1
2012–2015	−6.3	−3.6	6.6	6.4	7.4	6.6	0.1	3.8	13.1
2015–2018	−13.8	−13.1	−21.5	−3.1	−4.4	−4.7	−16.9	−17.5	−26.2
Gap between Private and Public schools									
2009–2012	5.1	8.6	8.6	−6.5	−6.0	−6.7	−1.4	2.6	1.9
2012–2015	−7.7	2.4	3.8	2.2	3.0	2.5	−5.6	5.5	6.3
2015–2018	−3.3	−10.9	−8.7	−11.3	−12.5	−12.3	−14.6	−23.5	−20.9

Conclusion

PISA is one of the most important instruments for policy assessment in education. A growing number of countries participate in the test and the release of the results has become an important milestone for researchers and policy makers. PISA is based on a sample designed to be representative of the country's 15 year-old student population. The private or public nature of the participating schools is one of the most widely used stratification variable used in PISA, and its results are often decomposed to compare student achievement by type of school.

The weight of the private system varies a lot across countries, and in some cases, across time, and the same happens regarding the weight of the vocational track. The analysis of PISA data often focus on inequalities and results disparities between students enrolled in different types of schools and school tracks.

In 2012 and 2015 the Portuguese results for the students enrolled in private schools showed a large increase, followed by a substantial fall in 2018. This contrasted with the evolution for students in public schools and with the PISA OECD average, spurring a discussion about the composition of the PISA sample in these years.

We compared the PISA sample distributions with the ones in the official student population data and found significant differences. In particular, in the PISA sample, we found a marked smaller share of the students in private schools in 2015, which is not observed in the 2018 wave. Additionally, in 2015, we observe a higher share of students in the academic track of the upper secondary compared with the ones enrolled in the lower secondary or in vocational tracks. In the 2018 sample both public and private schools subsamples are more aligned with the student population targeted by the OECD survey.

Using the true student population distribution, we make an adjustment to the official PISA scores. In 2015, for the students enrolled in private schools the adjusted scores are noticeably lower than those reported by OECD, while the adjustment for the scores in public schools is less pronounced. In 2018 the adjusted PISA scores are closer to the official results, mirroring the smaller difference found between the PISA sample and the official student population distribution. Thus, a substantial part of the fall observed in the PISA 2018 scores for the students in private schools can be accounted for a sizeable difference between the sample and population distributions in the previous PISA wave, in 2015. The use of adjusted sampling weights also changes the evolution of the private-public schools gap. The adjusted scores show that the gaps between the students in the two types of schools have shrunk from 2009, but in a smoother pattern and driven by the fall in the gap among students in the vocational track and by the growth of the share of students enrolled in this track in private schools.

Although the post-stratification approach used in this work yields an adjustment of the point estimates, it cannot address poor precision of these estimates and large confidence intervals when analysing specific subgroups for which sample sizes are small, as in the case of private schools in our study. A crucial takeaway from this is that when there are specific subgroups of interest, using bigger sample sizes and more effective designs is advised.

We also performed a decomposition of the evolution of PISA scores to determine if the variation across different PISA waves is explained by changes in the student population composition (population effect) or in the average student performance (score effect). Until the PISA wave in 2015 the decomposition shows similar results between private and public schools. Between 2015 and 2018 we find a negative contribution of the score and population effect for the students enrolled in the private system.

Achievement gaps between students from different subgroups, such as the ones enrolled in academic or vocational tracks or in public or private schools, are commonly viewed as indicators of inequality or of differences in the efficiency of education services. The observed differences between the distribution of the PISA sample and the Portuguese student population can emerge from several reasons, namely an incorrect initial sampling frame or not at random non-response rates across schools or of students within schools. PISA tries to control for non-response patterns in the weighting correction methods applied. However, a potential future disclosure by PISA of the initial sampling frame and the non-response patterns by school or student characteristics would improve the information about the PISA sample construction and shed light on eventual large differences relative to the target population. Additionally, when participating countries in PISA undergo sizeable changes in the share of students enrolled in public and private schools or reveal significant variations in achievement gap between different subgroups, an *ex-post* analysis, like the one used in this paper, could be a relevant tool to scrutinize and to better understand the sources of any such fluctuations in the PISA results.

Appendices

Appendix 1

Table 7

Table 7 Unweighted response rate

	2015		2018	
	School level	**Student Level**	**School level**	**Student Level**
Before school replacements	83.9%	–	85.2%	–
After school replacements	93.7%	82.2%	91.4%	76.1%

Appendix 2

Table 8

Table 8 Explicit and Implicit Stratification Variables

PISA years	Explicit stratification	Implicit stratification
2018	Geographic region (25); Certainty selections	ISCED (3); Funding (2); Urbanisation (3); Curriculum (3)
2015	Geographic region (25); Modal grade (2)	Funding (2); Urbanisation (3); ISCED programme orientation (3)

Appendix 3

Table 9

Table 9 Distribution PISA/Population students, with a two-year lag on Population data (year used for sampling)

Grade/Track of studies	2013			2018		
	Population	**PISA**	**Difference**	**Population**	**PISA**	**Difference**
All schools						
7th grade (Academic)	3.61%	2.65% (0.31)	−0.95[c]	3.31%	2.24% (0.26)	−1.07[c]
8th grade (Academic)	8.65%	7.04% (0.40)	−1.61[c]	7.92%	6.98% (0.46)	−0.94[b]
9th grade (Academic)	19.73%	19.88% (0.88)	0.15	17.94%	17.30% (0.96)	−0.64
Upper SE (Academic)	51.37%	57.18% (1.38)	5.81[c]	53.34%	57.50% (1.42)	4.16[c]
Lower SE (Vocational)	6.61%	4.60% (0.41)	−2[c]	6.89%	3.86% (0.87)	−3.03[c]
Upper SE (Vocational)	10.04%	8.65% (1.17)	−1.4	10.60%	12.11% (1.22)	1.51
Public schools						
7th grade (Academic)	3.96%	2.84% (0.33)	−1.13[c]	3.68%	2.56% (0.30)	−1.12[c]
8th grade (Academic)	9.32%	7.53% (0.40)	−1.79[c]	8.56%	7.97% (0.52)	−0.6
9th grade (Academic)	20.78%	20.81% (0.96)	0.03	19.37%	19.46% (1.00)	0.09
Upper SE (Academic)	51.35%	56.61% (1.43)	5.26[c]	53.82%	59.83% (1.31)	6.01[c]
Lower SE (Vocational)	7.25%	4.92% (0.43)	−2.32[c]	6.70%	2.36% (0.50)	−4.34[c]
Upper SE (Vocational)	7.34%	7.28% (0.76)	−0.06	7.87%	7.83% (0.76)	−0.04
Private schools						
7th grade (Academic)	1.24%	0.00%	−1.24	1.13%	0.39% (0.19)	−0.75[c]
8th grade (Academic)	4.16%	0.00%	−4.16	4.18%	1.36% (0.49)	−2.82[c]
9th grade (Academic)	12.78%	6.69% (4.71)	−6.09	9.69%	5.00% (1.19)	−4.69[c]
Upper SE (Academic)	51.49%	65.25% (10.89)	13.76	50.58%	44.23% (7.07)	−6.35
Lower SE (Vocational)	2.36%	0.00%	−2.36	8.01%	12.43% (4.65)	4.42
Upper SE (Vocational)	27.97%	28.06% (11.53)	0.09	26.40%	36.59% (6.22)	10.19

Statistically significant at [a]10%, [b]5% and [c]1% for one sample z-test of proportions. Standard errors for the PISA sample proportions are shown in parentheses next to the proportions (for private schools in 2015, three subgroups were not sampled). For each combination of year (2015, 2018) and type of school (All, Public, Private) we additionally performed chi-squared goodness-of-fit tests adjusted for survey design. Chi-squared statistics (p-values in parentheses): 2015 All schools: 112.94 (0.00); 2015 Private schools: 43.33 (0.24); 2015 Public schools: 103.32 (0.00); 2018 All schools: 130.59 (0.00); 2018 Private schools: 89.53(0.06); 2018 Public schools: 185.28 (0.00). Chi-square statistics calculated using the mgof STATA command, Jann (2008)

Appendix 4

In Table 2 we show that PISA did not sample any student enrolled in private schools in 2015 in the 7th and 8th grades (academic) and in the vocational lower secondary education. PISA scores for these subgroups of students were imputed, considering:

$$PISA^{2015}_{k,Private} = PISA^{2015}_{k,Public}\frac{PISA^{2015}_{\text{Upper secondary academic},Private}}{PISA^{2015}_{\text{Upper secondary academic},Public}}, \textit{ where } k = 7, 8 \text{ academic} \quad (3)$$

$$PISA^{2015}_{k,Private} = PISA^{2015}_{k,Public} \frac{PISA^{2015}_{\text{Upper secondary vocational},Private}}{PISA^{2015}_{\text{Upper secondary vocational},Public}}, \textit{ where } k = \text{Lower secondary vocational} \quad (4)$$

Appendix 5

See Table 10.

Table 10 Estimated proportion of students in private schools, by OECD country (weighted)

	2006	2009	2012	2015	2018
Australia	0%	40%	41%	39%	40%
Austria	9%	12%	9%	12%	12%
Belgium	68%	69%	67%	22%	0%
Canada	7%	7%	8%	8%	8%
Chile		53%	62%	60%	60%
Colombia					18%
Czechia	7%	3%	7%	8%	6%
Denmark	20%	23%	22%	19%	23%
Estonia		3%	4%	4%	4%
Finland	3%	4%	3%	4%	4%
France	0%	0%	18%	19%	18%
Germany	5%	5%	6%	6%	3%
Greece	5%	5%	6%	5%	5%
Hungary	16%	13%	16%	17%	20%
Iceland	1%	1%	1%	1%	1%
Ireland	60%	61%	58%	57%	0%
Israel		17%	0%	0%	0%
Italy	4%	6%	5%	3%	4%
Japan	31%	29%	30%	32%	34%
Latvia				2%	1%
Lithuania					4%
Luxembourg	14%	15%	15%	16%	18%
Mexico	15%	11%	12%	12%	12%
Netherlands	68%	65%	60%	39%	60%
New Zealand	6%	6%	5%	5%	6%
Norway	2%	1%	2%	2%	0%
Poland	2%	2%	3%	3%	4%
Portugal	10%	14%	10%	5%	13%
Slovakia	8%	9%	9%	12%	12%
Slovenia		3%	2%	2%	2%
South Korea	46%	37%	47%	35%	39%
Spain	35%	33%	33%	31%	32%

Table 10 (continued)

	2006	2009	2012	2015	2018
Sweden	8%	10%	14%	0%	0%
Switzerland	5%	6%	6%	6%	4%
Turkey	2%	1%	1%	5%	12%
United Kingdom	7%	6%	42%	5%	55%
United States	8%	9%	7%	8%	6%

Appendix 6

Table 11

Table 11 Estimated proportion of students for which public-private classification is missing, by OECD country (weighted)

	2006	2009	2012	2015	2018
Australia	100%	0%	0%	12%	4%
Austria	0%	3%	0%	1%	1%
Belgium	1%	0%	1%	64%	100%
Canada	3%	0%	0%	15%	1%
Chile		9%	2%	3%	8%
Colombia					1%
Czechia	2%	5%	14%	3%	3%
Denmark	17%	1%	8%	22%	20%
Estonia		0%	0%	1%	0%
Finland	0%	0%	1%	3%	1%
France	100%	100%	5%	8%	11%
Germany	4%	6%	13%	23%	14%
Greece	1%	0%	0%	1%	2%
Hungary	4%	0%	1%	7%	1%
Iceland	4%	9%	1%	3%	2%
Ireland	1%	0%	0%	0%	100%
Israel		3%	2%	100%	100%
Italy	1%	2%	6%	28%	2%
Japan	0%	0%	0%	0%	0%
Latvia				1%	2%
Lithuania					0%
Luxembourg	0%	0%	0%	0%	0%
Mexico	1%	0%	0%	0%	1%
Netherlands	0%	2%	10%	35%	5%
New Zealand	0%	0%	11%	21%	0%
Norway	2%	2%	5%	17%	100%
Poland	0%	0%	1%	1%	0%
Portugal	0%	0%	1%	1%	0%
Slovakia	1%	0%	0%	0%	5%
Slovenia		0%	5%	8%	8%
South Korea	0%	0%	0%	0%	1%
Spain	0%	2%	0%	0%	3%
Sweden	1%	0%	0%	100%	100%
Switzerland	1%	1%	2%	7%	1%
Turkey	1%	0%	0%	1%	0%
United Kingdom	8%	5%	5%	24%	14%
United States	2%	0%	2%	2%	9%

Appendix 7

See Table 12.

Table 12 Estimated proportion of students in vocational tracks, by OECD country (weighted)

	2006	2009	2012	2015	2018
Australia	11%	13%	11%	13%	10%
Austria	54%	54%	60%	67%	58%
Belgium	36%	47%	42%	32%	42%
Canada	0%	0%	0%	0%	0%
Chile		2%	3%	1%	2%
Colombia					19%
Czechia	41%	36%	31%	33%	33%
Denmark	0%	1%	0%	1%	1%
Estonia		0%	0%	0%	0%
Finland	0%	0%	0%	0%	0%
France	9%	0%	0%	0%	0%
Germany	57%	52%	51%	50%	45%
Greece	14%	14%	13%	16%	13%
Hungary	57%	51%	50%	51%	47%
Iceland	0%	0%	0%	0%	0%
Ireland	4%	2%	5%	4%	3%
Israel		0%	0%	0%	0%
Italy	56%	55%	50%	50%	49%
Japan	26%	25%	24%	24%	23%
Latvia				1%	1%
Lithuania					2%
Luxembourg	13%	14%	15%	51%	50%
Mexico	31%	34%	34%	34%	33%
Netherlands	78%	78%	75%	71%	78%
New Zealand	0%	0%	0%	0%	0%
Norway	0%	0%	0%	0%	0%
Poland	0%	0%	0%	0%	0%
Portugal	12%	9%	7%	13%	16%
Slovakia	19%	18%	9%	8%	8%
Slovenia		61%	61%	64%	63%
South Korea	0%	0%	0%	16%	16%
Spain	0%	0%	1%	1%	1%
Sweden	1%	1%	0%	0%	0%
Switzerland	10%	10%	13%	12%	14%
Turkey	41%	36%	34%	41%	33%
United Kingdom	0%	0%	0%	0%	0%
United States	0%	0%	0%	0%	0%

Acknowledgements
The authors thank the editor and the anonymous referees for their valuable comments. We also thank DGEEC for providing access to the Portuguese education data.

Author contributions
All authors developed the concept of the study. PF and RC wrote the first draft of the manuscript, along with undertaking the analysis. All authors contributed to the editing of the paper and the development of the final manuscript. All authors have read and approved the final manuscript.

Funding
This work received supported from Fundação para a Ciência e a Tecnologia, in Portugal (PTDC/EGE-ECO/4764/2021, UIDB/00124/2020, UIDP/00124/2020, Social Sciences DataLab PINFRA/22209/2016 and POR Lisboa and POR Norte; Ricardo Colaço's scholarship PD/BD/150428/2019).

Availability of data and materials
The PISA datasets analysed in this study are available at https://www.oecd.org/pisa/data/ and the Portuguese administrative data is available for research purposes upon request to DGEEC/Portuguese Ministry of Education.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All the authors consented for publication.

Competing interests
The authors declare that they have no Conflict of interest.

Received: 5 June 2023 Accepted: 3 December 2024
Published online: 13 January 2025

References

Anders, J., Has, S., Jerrim, J., Shure, N., & Zieger, L. (2020). Is Canada really an education superpower? The impact of non-participation on results from PISA 2015. *Evaluation and Accountability Educational Assessment., 33*, 229–49.

Andersson, C., & Massih, S. S. (2023). Pisa 2018: did Sweden exclude students according to the rules? *Assessment in Education: Principles, Policy & Practice., 30*(1), 33–52.

Avvisati, F., & Keslair, F. (2016). *REPEST: Stata module to run estimations with weighted replicate samples and plausible values*.

Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2018). Testing NBER Working Paper No 24836. *National Bureau of Economic Research*. 59:349-88.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources, 8*, 436.

Delprato, M., & Chudgar, A. (2018). Factors associated with private-public school performance: analysis of TALIS-PISA link data. *International Journal of Educational Development, 61*, 155–72.

Durrant, G. B., & Schnepf, S. V. (2018). Which schools and pupils respond to educational achievement surveys?: a focus on the English Programme for International Student Assessment sample. *Journal of the Royal Statistical Society Series A: Statistics in Society, 181*(4), 1057–75.

Freitas, P., Nunes, L. C., Balcão Reis, A., Seabra, C., & Ferro, A. (2016). Correcting for sample problems in PISA and the improvement in Portuguese students' performance. *Assessment in Education: Principles, Policy and Practice, 23*(4), 456–472.

Gevrek, Z. E., Gevrek, D., & Neumeier, C. (2020). Explaining the gender gaps in mathematics achievement and attitudes: the role of societal gender equality. *Economics of Education Review, 76*, 101978.

Heine, J. H., Nagy, G., Meinck, S., Zühlke, O., & Mang, J. (2017). Empirische grundlage, stichprobenausfall und adjustierung im pisa-längsschnitt 2012–2013. *Zeitschrift fur Erziehungswissenschaft, 2*(20), 287–306.

Jann, B. (2008). Multinomial goodness-of-fit: large-sample tests with survey design correction and exact tests for small samples. *Stata Journal, 8*, 147–69.

Jerrim, J. (2013). The reliability of trends over time in international education test scores: is the performance of England's secondary school pupils really in relative decline? *Journal of Social Policy, 42*(2), 259–79.

Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK? *Review of Education, 9*(3), e3270.

Lavy, V. (2010). Do differences in school's instruction time explain International Achievement gaps in math, science, and reading? Evidence from developed and developing countries. NBER Working Paper No. 16227. *National Bureau of Economic Research*.

Mancebón, M. J., Calero, J., Choi, Á., & Ximénez-De-Embún, D. P. (2012). The efficiency of public and publicly subsidized high schools in Spain: evidence from PISA-2006. *Journal of the Operational Research Society, 63*(11), 1516–33.

Micklewright, J., Schnepf, S. V., & Skinner, C. (2012). Non-response biases in surveys of schoolchildren: The case of the english programme for international student assessment (pisa) samples. *Journal of the Royal Statistical Society Series A: Statistics in Society, 175*, 915–38.

Neuwirth, E. (2006). PISA 2000: Sample weight problems in Austria. *OECD EducationWorking Papers, No. 5*.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review, 14*, 693.

OECD (2009). *PISA data analysis manual: SPSS, Second Edition*.
OECD (2017). Pisa 2015 technical report. *OECD Publishing*.
OECD (2020). *PISA 2018 Results (Volume V): effective policies, successful schools*, volume V.
O'Leary, M. (2001). The effects of age-based and grade-based sampling on the relative standing of countries in international comparative studies of student achievement. *British Educational Research Journal, 27*(2), 187–200.
Perera, L. D. H., & Asadullah, M. N. (2019). Mind the gap: what explains Malaysia's underperformance in Pisa? *International Journal of Educational Development, 65*, 254–63.
Vandenberghe, V., & Robin, S. (2004). Evaluating the effectiveness of private education across countries: a comparison of methods. *Labour Economics, 11*(4), 487–506.
Woessmann, L. (2016). The importance of school systems: evidence from international differences in student achievement. *In Journal of Economic Perspectives, 30*, 3–31.

Publisher's Note